AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 105:1–7 (1998)

Mitochondrial DNA and Ancient Population Growth

JOHN H. RELETHFORD*
Department of Anthropology, State University of New York College at Oneonta, Oneonta, New York 13820

KEY WORDS mitochondrial DNA; population growth; mismatch distributions; modern human origins

ABSTRACT In recent years, the study of mitochondrial DNA (mtDNA) variation has entered a new phase with an increasing emphasis on interpretations of demographic, rather than phylogenetic, history. Human mtDNA variation fits a "sudden expansion" model, where the human species expanded rapidly in size during the Late Pleistocene. This paper examines the sudden expansion model with the goal of partitioning total mtDNA diversity in contemporary populations into two components—diversity that existed prior to the population expansion and diversity that arose after the expansion. A method is developed for estimating these components. Analysis of mtDNA diversity within selected human populations shows that 64–80% of mtDNA diversity in contemporary populations arose after the expansion, a consequence of a high mutation rate relative to the number of generations since expansion. The basic model is extended to two components of excess diversity in sub-Saharan Africa—differences in population size before the expansion and differences in the timing of expansion. Results suggest that excess sub-Saharan African mtDNA diversity is due to the combined effects of the sub-Saharan African population being larger in size prior to the expansion *and* expanding earlier. Am J Phys Anthropol 105:1–7, 1998.

The application of mitochondrial DNA (mtDNA) to questions of modern human origins has had a short and controversial history. From an anthropological perspective, the key event was the paper by Cann et al. (1987) which claimed support for a recent (<200 kyr) and singular African origin of modern humans. This replacement model contrasts with the multiregional model of human evolution, which proposes that modern humans arose from evolutionary changes within a single evolutionary lineage dating back to the origin of *Homo erectus* roughly two million years ago. The support for a replacement model was based on two findings from the mtDNA analysis—deep African roots in the gene tree and the higher sequence diversity in sub-Saharan African populations. These findings were later replicated with a larger data set by Vigilant et al. (1991). Doubts then arose over the phylogenetic analysis because of problems with the initial analysis (Hedges et al., 1992; Templeton, 1992). The issue of excess African diversity also came into question with the demonstration that diversity is not likely to track the age of a population (Rogers and Jorde, 1995), and with analyses showing a larger long-term effective population size in Africa (Relethford and Harpending, 1994; Relethford, 1995).

Another approach to mtDNA variation, developed in the past few years, focuses on patterns of variation as clues to the *demo-*

*Correspondence to: Dr. John Relethford, Department of Anthropology, State University of New York College at Oneonta, Oneonta, NY 13820. Fax: (607) 436-2653.
E-mail: relethjh@oneonta.edu

Received 11 February 1997; accepted 20 October 1997.

graphic, rather than *phylogenetic,* history of our species. This work has demonstrated that there was a dramatic increase in the effective population size of the human species in the Late Pleistocene (Rogers and Harpending, 1992; Harpending et al., 1993; Sherry et al., 1994; Rogers, 1995, 1997; Rogers and Jorde, 1995). Related work has demonstrated that long-term effective population size was larger in sub-Saharan Africa (Relethford and Harpending, 1994), and that variation in ancient population size can mimic the genetic effects expected under a recent African origin model (Relethford and Harpending, 1994, 1995). There is now a growing tendency to view genetic data, particularly mtDNA, as providing clues to ancient demography, rather than a direct phylogenetic statement (Relethford, 1995; Harpending et al., 1996).

The purpose of this paper is to examine two ways in which to extract additional demographic information from mtDNA variation. Both of these approaches build upon the demonstration by Rogers and colleagues of a Late Pleistocene population explosion where small human populations expanded in size rapidly during the past 100,000 years. The first approach provides a way to partition modern mtDNA sequence diversity into two components—diversity that arose *before* the expansion and diversity that arose *after* the expansion. The second approach is an extension to the comparison between two populations, and addresses contributions to excess mtDNA diversity in sub-Saharan populations.

THE MODEL OF SUDDEN EXPANSION

The sudden expansion model derives from a consideration of the properties of the mitochondrial DNA mismatch distribution, which is a histogram of the number of pairwise differences in mtDNA sequences. Each individual in a sample is compared to each other individual to determine the number of nucleotide or restriction site differences between them. The mismatch distribution of most human populations shows a smooth wave-shaped curve. Rogers and Harpending (1992) found that such curves do not resemble the curves expected under a theoretical model that assumes an equilibrium distribution obtained from a constant population size over time. Instead, they showed that the observed curves reflect a rapid and extensive ancient population explosion, with the peak of the mismatch distribution corresponding to the time of a past population expansion. It is important to remember that *expansion* in this model refers to an expansion of population *size,* and not necessarily an expansion of the geographic distribution of a population.

The Rogers-Harpending model can be expressed in terms of a simple model of demographic change. The model assumes an initial population of size N_0 that increases instantaneously to size N_1 at t generations in the past (as applied to mtDNA, all values of N refer to the effective number of females in the population). Changes in equilibrium mtDNA sequence diversity (the number of nucleotide/restriction site differences between pairs of individuals) are directly proportional to changes in population size. That is,

$$\begin{aligned} \theta_0 &= 2\mu N_0 \\ \theta_1 &= 2\mu N_1 \end{aligned} \qquad (1)$$

where θ_0 is the equilibrium mtDNA sequence diversity prior to population expansion, θ_1 is the equilibrium mtDNA sequence diversity after population expansion, and μ is the aggregate mutation rate over all sites. The θ values can be thought of as population sizes expressed in terms of mutational units (1/2μ females). The time of population expansion in generations (t) can also be expressed in terms of mutational units (1/2μ generations) as

$$\tau = 2\mu t \qquad (2)$$

Under this model, the mtDNA sequence diversity is assumed to be at equilibrium prior to expansion (θ_0) with a small initial inbreeding effective female population size (N_0). This is a reasonable assumption since mtDNA diversity (and most genetic measures) converge rapidly to equilibrium when population size is small (Rogers and Harpending, 1992; Rogers and Jorde, 1995). The population then increases instantaneously t generations in the past (τ mutational units in the past) to a much larger size

N_1, and the population then begins moving toward the new equilibrium diversity θ_1.

While the initial development of the sudden expansion model proposed methods to estimate three parameters (θ_0, θ_1, and τ), better results were obtained by setting N_1 to infinity, which provides a simple approximation to the case where N_1 is merely large relative to N_0 (Rogers, 1995). The simplified two-parameter model provides estimates of pre-expansion mtDNA diversity (θ_0) and mutational time since expansion (τ). Given an appropriate estimate of the mutation rate μ, these estimates can be used with equations (1) and (2) to estimate pre-expansion size (N_0) and the number of generations since expansion (t). Given an estimate of generation length, t can also be converted into years.

While mathematically interesting, simplified models are useful only if their assumptions are valid and they show a good fit to real data. To date, the model has proven remarkably robust to violations of the assumptions, including starting from equilibrium θ_0 (Rogers and Jorde, 1995) and use of an infinite-sites mutation model (Rogers, 1992; Rogers et al., 1996). The assumption of instantaneous population growth, needed for model simplification, has been shown to approximate closely more complex demographic histories and trajectories of population growth (Rogers and Harpending, 1992), including cases where there is a series of population expansions and bottlenecks (Rogers, 1997). Different variations of population subdivision do not bias results (Rogers, 1997). The model works well with low-resolution, as well as high-resolution, mtDNA data (Harpending, 1994).

The two-parameter model has now been applied to mtDNA data from 25 human samples from around the world. Of these, 23 show an excellent fit to the model of sudden expansion, and the two outliers are cases with known recent population bottlenecks that compromise results (Sherry et al., 1994). Several analyses have estimated a total species size, prior to expansion, of roughly 1,500 to 10,000 females (for a total species effective size of 3,000 to 20,000 individuals) (Harpending et al., 1993; Rogers, 1995, 1997; Rogers and Jorde, 1995). The long-term effective species size will be larger than the pre-expansion size, so that a rough mid-range estimate of long-term effective size is on the order of 10,000. It is important to note that such estimates will vary depending on the mutation rate used, which in turn is dependent on the calibration point used, most often the divergence of chimpanzees and humans. Nonetheless, the rough mid-range species effective size estimate of roughly 10,000 is in agreement with other genetic studies (e.g., Klein et al., 1993; Takahata, 1993).

Simulation studies suggest that the population explosion was quite large, with increases of at least 100-fold (Rogers, 1995). Estimates of expansion time are consistently less than 100,000 years, with an average expansion time of roughly 40,000 to 60,000 years (Sherry et al., 1994). There is a suggestion that expansion was earlier in sub-Saharan African populations (Sherry et al., 1994), which ties in with the finding of a larger long-term effective population size in sub-Saharan Africa (Relethford, 1995). Recent work suggests that the small pre-expansion species size may have resulted from a bottleneck estimated to have occurred roughly 95,000 to 130,000 years ago (Sherry, 1996).

DIVERSITY BEFORE AND AFTER EXPANSION

Given this brief review, it is useful to consider several implications for the study of the origins of modern human genetic diversity. Taking mtDNA diversity as a reflection of history we can ask questions regarding the relative impact of prior demographic events, particularly small pre-expansion size and the later population explosion. In terms of the sudden expansion model, how much of our contemporary mtDNA diversity is due to diversity that accumulated *prior* to the expansion, and how much is due to the accumulation of diversity *after* the expansion (Harpending et al., 1996)?

To answer this, we must consider the evolution of mtDNA diversity over time. In terms of the sudden expansion model, the

sequence diversity observed t generations after expansion, $\theta(t)$, can be written as

$$\theta(t) = \theta_1 + (\theta_0 - \theta_1)e^{-t/N_1} \quad (3)$$

(Li, 1977; Rogers and Jorde, 1995). By definition, θ_0 is the amount of diversity in the population prior to expansion. Let θ_a represent the amount of diversity that accumulates *after* the expansion. That is, $\theta_a = \theta(t) - \theta_0$, which is equal to

$$\theta_a = (\theta_1 - \theta_0)\,(1 - e^{-t/N_1}) \quad (4)$$

If N_1 is very large relative to t, the term e^{-t/N_1} can be approximated as $(1 - t/N_1)$, so that equation (4) can be approximated as

$$\theta_a \approx \frac{(\theta_1 - \theta_0)t}{N_1} \quad (5)$$

Substituting equations (1) and (2) into equation (5) gives

$$\theta_a \approx 2\mu t\left(1 - \frac{N_0}{N_1}\right) \quad (6)$$

Since the model states that N_0 is very small relative to N_1, and since Rogers (1995) has shown that the ratio N_0/N_1 is less than 0.01, equation (6) can be further approximated as $\theta_a \approx 2\,\mu t$, which by definition (equation 2) means that $\theta_a \approx \tau$. As such, observed mtDNA diversity can be expressed in terms of a simple approximation using the parameters of the sudden expansion model as

$$\theta(t) \approx \theta_0 + \tau \quad (7)$$

(see also Harpending, 1994; Rogers, 1995). The first term in equation (7) estimates diversity accumulated prior to the expansion, and depends on initial population size. The second term in equation (7) estimates diversity accumulated after the expansion and depends on the number of generations since the expansion. Both parameters can be estimated from the observed mismatch distribution as

$$\hat{\theta}_0 = \sqrt{v - m}$$
$$\hat{\tau} = m - \hat{\theta}_0 \quad (8)$$

where m and v are the observed mean and variance of the mismatch distribution (Harpending et al., 1993; Rogers, 1995). As shown in equations (1) and (2), these parameters represent initial population size and time since expansion as expressed in mutational units, which can be converted given estimates of the mutation rate. Equation (7) also shows that these parameters provide valuable information on the relative contribution of pre-expansion and post-expansion histories to contemporary diversity. Further, estimates of these relative contributions ($\theta_0/[\theta_0 + \tau]$ and $\tau/[\theta_0 + \tau]$) do not require estimation of the mutation rate.

Equation (7) also shows that when initial population size (N_0) is equal to the number of generations since expansion (t) then the relative contributions are equal. Further, when $N_0 > t$, then there is a greater relative contribution prior to expansion, and when $N_0 < t$, then there is a greater relative contribution after expansion.

Table 1 provides estimates of these relative contributions for several selected human samples using estimates of θ_0 and τ reported by Sherry et al. (1994) based on restriction fragment length polymorphism (RFLP) data. These results show clearly that contemporary mtDNA diversity is the result of diversity accumulated both before and after population expansion, but that in all cases the majority of diversity (64–80%) accumulated *after* expansion. A greater post-expansion contribution is expected given the high mutation rate for mitochondrial DNA; the relative contributions would not be expected to be the same for traits with lower mutation rates. It is important to stress that these estimates do not have standard errors. The estimates of relative contributions to diversity must not be taken as exact, but rather as first-order approximations.

DIFFERENCES IN DIVERSITY BETWEEN POPULATIONS

Several studies have noted higher levels of mtDNA sequence diversity in sub-Saharan populations relative to other regions (Cann et al., 1987; Stoneking and Cann, 1989; Vigilant et al., 1991). Some authors have suggested that higher sub-Saharan African diversity reflects a greater age for African populations under a recent African origin model, since there has been more time for the accumulation of mutations. In this interpretation, diversity is a direct reflection

TABLE 1. Estimated parameters ($\hat{\theta}_0$, $\hat{\tau}$) of population expansion (from Sherry et al., 1994) and the percentages of total diversity accumulated before and after expansion based on RFLP data

Population[1]	$\hat{\theta}_0$	$\hat{\tau}$	Estimated expansion time (kyr)[2]	Percentage of diversity due to accumulation	
				Before expansion	After expansion
Africans (n = 20)	4.07	8.95	100	31	69
Europeans (n = 47)	1.70	4.99	56	25	75
Asians (n = 34)	1.97	7.88	88	20	80
Papua New Guinea (n = 21)	1.95	4.26	48	31	69
Australians (n = 21)	2.49	4.40	49	36	64
World sample (n = 241)	2.45	5.50	61	31	69

[1] Numbers in parentheses are sample sizes.
[2] Estimated expansion time in years is taken from Sherry et al. (1994).

of time since a population's origin. However, recent work has suggested that this interpretation makes assumptions that are not likely in human evolution (Relethford, 1995; Rogers and Jorde, 1995). An alternative interpretation is that Africa had the largest long-term effective population size (Relethford and Harpending, 1994, 1995; Relethford, 1995), an interpretation compatible with both replacement and multiregional models (Wolpoff and Caspari, 1997).

An effective population size is useful mathematically, but can mask a variety of patterns of demographic change. Possible models are limited by the sudden expansion model—rapid growth from a small to a very large population. Given this general model, one population could have a larger long-term effective population size if it was initially larger and/or it expanded earlier (Relethford and Harpending, 1994; Relethford, 1995; Harpending et al., 1996). Was Africa larger prior to expansion? Did Africa expand earlier than other regions? These questions can be addressed by considering the separation of total diversity into pre-expansion and post-expansion components.

Consider two populations, A and B, with contemporary sequence diversities of θ_A and θ_B respectively. Assume that both populations fit the sudden expansion model, but with different values of pre-expansion diversity (θ_{0_A}, θ_{0_B}) and time since expansion (τ_A, τ_B). Let Δ represent the excess diversity in population A; that is,

$$\Delta = \theta_A - \theta_B \tag{9}$$

Substituting equation (7) gives the approximation

$$\Delta \approx (\theta_{0_A} - \theta_{0_B}) + (\tau_A - \tau_B) \tag{10}$$

The first half of equation (10) represents the amount of excess diversity due to initial differences in diversity, and hence population size. The second half of equation (10) represents the amount of excess diversity due to differences in the timing of population expansion. It is clear from equation (10) that if the population sizes of A and B are equal then the excess diversity of A is due entirely to differences in the timing of expansion. Likewise, if populations A and B both expanded at the same time then the excess diversity of A is entirely due to differences in initial population size.

Data from Table 1 were used to estimate the two components that contribute to excess African diversity using equation (10), and the results are presented in Table 2. Roughly 70% of excess African diversity is due to an earlier African expansion relative to Europe, New Guinea, and Australia, and about 30% due to a larger pre-expansion African population size. Comparing African and Asian diversity, however, shows a greater contribution (66%) due to initial differences in population size, and less (34%) due to differences in the estimated timing of population expansion. These differences relate directly to the parameters reported in Table 1; Africa expanded first, followed closely by Asia, and followed later by Europe and Australasia. Despite the differences among comparisons in Table 2, the general trend is clear—differences in contemporary mtDNA

TABLE 2. Components of excess African diversity for selected regional comparisons based on RFLP data

Comparison	Percentage of excess African diversity due to: Larger pre-expansion African population size	Percentage of excess African diversity due to: Earlier African expansion
Africans/Europeans	37	63
Africans/Asians	66	34
Africans/Papua New Guinea	31	69
Africans/Australians	26	74

diversity appear to reflect *both* differences in pre-expansion population size *and* differences in the timing of population expansion.

DISCUSSION

Given issues of sample size and composition, reliance on a single genetic trait, and a lack of standard errors, we should not read too much into the *specific* estimates presented in Tables 1 and 2. The general patterns, however, are clear. The mitochondrial DNA diversity found within contemporary human populations appears to have accumulated *both* before *and* after population expansion. Excess African mtDNA diversity is also a function of pre-expansion and post-expansion demographic history; specifically, Africa was larger before the expansion, and also expanded earlier relative to other regions.

These findings tie together previous studies demonstrating a Late Pleistocene "population explosion" (Rogers, 1995, among others) and the finding of a larger long-term effective population size in sub-Saharan Africa (Relethford and Harpending, 1994). There are three different ways in which Africa could have a larger long-term effective population size under the sudden expansion model: 1) Africa was larger prior to an expansion which took place across the Old World at the same time; 2) there was no difference in population size among geographic regions across the Old World prior to the expansion, but Africa expanded earlier than other regions; or 3) Africa was larger prior to the expansions, *and* it also expanded earlier. The results presented here favor the third hypothesis, since there seems to be a non-trivial contribution of both initial size and timing of expansion towards excess African diversity.

These findings do not bear directly on arguments over modern human origins. While variation in population size and timing of expansion have been discussed in terms of a replacement model (e.g., Harpending et al., 1993), they are also compatible with the multiregional model, which predicts a larger long-term effective African population and a larger pre-expansion African population (Wolpoff and Caspari, 1997).

Limitations of these analyses must be kept in mind, particularly the lack of standard errors and the reliance on the equivalent of a single genetic locus. Therefore, the results should be taken as suggestive, and not conclusive. It remains to be seen whether such methods can be extended to other sorts of genetic data, and whether similar results can be found. Nonetheless, there is sufficient suggestion of differences in both pre-expansion size and timing of expansion to warrant attention at a paleoanthropological and archaeological level. In particular, can we find archaeological or paleoenvironmental data supporting a larger pre-expansion African population? Can the hypothesis of earlier African expansion be supported by archaeological data? Does expansion in population *size* relate to a possible expansion in population *distribution?* These, and other, questions must ultimately be addressed by a synthesis of genetic, fossil, and archaeological data. The results presented here do not bear directly on the question of multiregional evolution versus replacement, but instead refocus the genetic data on questions of ancient demography.

ACKNOWLEDGMENTS

I thank Henry Harpending and Milford Wolpoff for their comments.

LITERATURE CITED

Cann RL, Stoneking M, and Wilson AC (1987) Mitochondrial DNA and human evolution. Nature *325:*31–36.

Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum. Biol. *66:*591–600.

Harpending HC, Sherry ST, Rogers AR, and Stoneking M (1993) The genetic structure of ancient human populations. Curr. Anthropol. *34:*483–496.

Harpending HC, Relethford JH, and Sherry ST (1996) Methods and models for understanding human diversity. In AJ Boyce, and CGN Mascie-Taylor (eds.):

Molecular Biology and Human Diversity. Cambridge: Cambridge University Press, pp. 283–299.
Hedges SB, Kumar S, Tamura K, and Stoneking M (1992) Human origins and analysis of mitochondrial DNA sequences. Science *255:*737–739.
Klein J, Takahata N, and Ayala FJ (1993) MHC polymorphism and human origins. Sci. Am. *269(12):*78–83.
Li W-H (1977) Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. Genetics *85:*331–337.
Relethford JH (1995) Genetics and modern human origins. Evol. Anthropol. *4:*53–63.
Relethford JH, and Harpending HC (1994) Craniometric variation, genetic theory, and modern human origins. Am. J. Phys. Anthropol. *95:*249–270.
Relethford JH, and Harpending HC (1995) Ancient differences in population size can mimic a recent African origin of modern humans. Curr. Anthropol. *36:*667–674.
Rogers AR (1992) Error introduced by the infinite-sites model. Mol. Biol. Evol. *9:*1181–1184.
Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. Evolution *49:*608–615.
Rogers AR (1997) Population structure and modern human origins. In PJ Donnelly and S Tavaré (eds.): Progress in Population Genetics and Human Evolution. New York: Springer-Verlag, pp. 55–79.
Rogers AR, and Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. *9:*552–569.
Rogers AR, and Jorde LB (1995) Genetic evidence on modern human origins. Hum. Biol. *67:*1–36.
Rogers AR, Fraley AE, Bamshad MJ, Watkins WS, and Jorde LB (1996) Mitochondrial mismatch analysis is insensitive to the mutational process. Mol. Biol. Evol. *13:*895–902.
Sherry ST (1996) Estimating Human Effective Population Sizes with Genetic Models Incorporating Demographic Fluctuation. PhD dissertation, Department of Anthropology, Pennsylvania State University.
Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, and Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. Hum. Biol. *66:*761–775.
Stoneking M, and Cann RL (1989) African origin of human mitochondrial DNA. In P Mellars and C Stringer (eds.): The Human Revolution: Behavioral and Biological Perspectives of the Origin of Modern Humans. Princeton: Princeton University Press, pp. 17–30.
Takahata N (1993) Allelic genealogy and human evolution. Mol. Biol. Evol. *10:*2–22.
Templeton AR (1992) Human origins and analysis of mitochondrial DNA sequences. Science *255:*737.
Vigilant L, Stoneking M, Harpending H, Hawkes K, and Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science *253:*1503–1507.
Wolpoff M, and Caspari R (1997) Race and Human Evolution. New York: Simon and Schuster.